A model for human activities recognition using partially occluded images

Fadheela Sabri Abu-Almash
Ministry of Higher Education and Scientific Research
Baghdad, Iraq

Abstract:

The demand for Human Activities Recognition (HAR) from still images and corresponding features categorization using appropriate classifier is never ending. HAR of partially occluded object using still image is highly challenging than extracting information from video due to the absence of any prior knowledge resembling frames stream .This new research domain dealing with the computer identification and subsequent classification of specific activities to develop understanding of human behaviors has diversified applied interests in surveillance cameras, security systems and automotive industry. We propose a new model for HAR with occluded missing part of human body in still image.(A template is built to complete the full appearance predictable by the system and their subsequent features classification is made.)The problems of HAR in still images are addressed using nonlinear Support Vector Machine classifier. The process of de-noising and chamfer matching are performed. This model is simulated with1200 still images of (64 x 128) pixels based on existing datasets such (INRIA and KTH). Using this model ,a recognition rate 86% is achieved for seven activities such as running, walking, jumping, clapping, jogging, boxing and waving .The excellent features of the results suggest that our method may constitute a basis for HAR with occlusion human body parts in still images promising for accurate features classification.

Keywords: Human activities recognition, still image, occlusion human body, features extraction.

Fadheela .S

1. Introduction:

Lately, HAR using still images via the inference of human pose became the key issue in precisely characterizing and classifying the behavioral pattern .The major difficulties in estimating human postures from still image using partially occluded human body in the scene provided further impetus in features categorization due to its relevance towards sundry applications [1]. One of the most popular approaches in this research domain is the pictorial structure framework [2,3]. Despite many efforts, optimal and efficient models for HAR are far from being developed.

Detail examinations of the activities from still images when the human body is partially occluded in the scene are our interests. However, training the activities recognition system with only available knowledge on the rigid parts of the image without any other prior information is indeed a challenging task. The main excitement lies in extracting the full feature from partially occluded object and then classifying it for the human body. Unlike other methods [4] that use the entire body and their full feature matching ,the present approach is completely new. Undoubtedly, there is a need to establish an optimal model for recognizing human activities with high level of accuracy which may be accomplished by a statistical measure based on the data likelihood. Recently, intensive experimentations are conducted on several datasets[5] to enable an understanding on how many activities recognition are supported.

In the last decade, human activities recognition of video images is widely studied. Several methods are proposed to improve the accuracy and robustness to cope with challenging video .Many algorithms are developed to achieve accurate results with high recognition rate as much as 100% for single activity .The most important factor is the unavailability of prior knowledge where the generated current information can be compared with the previous one as a frame in video sequences .Although, the human activities recognition using still image being poor in terms of prior knowledge even then

good and powerful features are extracted from the human body [6].For effective and influential features the object should be integral part of the image .In this viewpoint we propose a method to complete the occluded object in still image then extract features and classify the activities to get them recognized.

The recognition of human interaction with other objects in realistic video demonstrated a correlation between the entities and behavior of the human parts [7].Based on this interaction pattern recognition, a method is developed to keep track of the hand part with corresponding distance of the head and this kind of feature is found to be suitable when some activities including drinking, smoking, eating, and so on are needed .This paper aimed to represent the object trajectory by drawing the path for each object to compare with the position of human part. This approach can efficiently reveal a relation between object and human parts. The occlusion of some parts of the human body does not cause any problem provided the activity must be related and compatible with those particular objects such as smoking and drinking.

Bo and Ram [8] introduced an approach to automatically detect and track multiple with possible partially occluded human in two activities of walking and standing pose given by one camera .Essentially, a human body is treated as an assembly of body parts. This method learnt by boosting a number of weak classifiers which are based on edge-let features. The occluded parts of the body are solved by joints likelihood model including an analysis of possible occlusions. Sung J. and Mark S. proposed another method [9] comprised of detecting the heel strike using gait trajectory model containing many striking features such as robustness to occasion, camera view and low resolution. This head movement based method is conspicuous and sinusoidal in which the region of interest is extracted as feature from silhouette image. The number of times the heel strike is regarded to find the trajectory model which is effective when used for HAR in video stream.

Fadheela .S

The detection of human behavior remains a challenge in computer vision and HAR especially due to highly articulated body postures, changes in viewpoints, varying illumination condition and clustered background. To avoid these difficulties most of the previous studies on HAR often considered on low-articulated postures. Trung and Phong [10] proposed a new method to detect human region from still images using raw edges. The object fixed in the still images is mostly obtained from pedestrians with varieties of occurrences inside the image .This approach produced better results than sliding window-style method for detecting human activities .The method relies on characteristics of boundaries and interest points that combine different image processing techniques such as image mean algorithm and probability for choosing a human region .INRIA challenged dataset is used because it possesses multi-activities with several view-points in addition to the occlusion clusters under varying length condition indoor and outdoor scenes.

Precisely, each HAR system must pass through several stages processing such as removal of noise, segmentation, feature extraction and classification. Mehmet and Alper established a technique composed of a fuzzy inference system and an edge detection and dilation unit for removing the noise from image to address the speckle noisy image [11]. An efficient and accurate feature selection and classification method is introduced by Habil [12] to achieve a superior classifier .This approach is capable of learning the system by running it simultaneously in an online manner to estimate the parameters within a nearest neighbor with Gaussian mixture model. Halime [13] developed a Cellular Neural Network model to calculate the adaptive iterative value via wavelet transform and spatial frequency for segmenting the mutable region inside an image .An integrated automatic image capture strategy for automatically analyzing large numbers of image via evaluation software is used .A learning based framework for action representation and recognition on the description of an action by time series of optical flow motion features is reported by Michalis and Vasileios [14].

The Gaussian mixture modeling is used to cluster the activities .Excellent experimental results are achieved by using Weizmann, KTH, UCF sports and UCF you tube action datasets.

In this paper, we address the problems of HAR in still image having partial occlusion using one of the most popular classifier called Support Vector Machine(SVM).High accuracy is achieved for the features classification when applied to the images having a missing part of the human body or occlusion .The model is simulated on INRIA and KTH datasets with several still images using Matlab programming. This work renders three major contributions .Firstly, the completion of the occluded object (a part of human body) that occurs out of the scenes when the object is hidden .The features are then extracted from this occluded part to suggest the remaining parts of the body. Secondly, the proposed new features are highly compatible with the nature of activities preferred for classification .Lastly ,a high recognition rate is achieved using the non-linear SVM classifier for seven activities such as walking, running, jogging, waving, jumping, clapping, and jumping.

2. Methodology

To achieve Human Activities Recognition system we take in our consideration five main stages first start with preprocessing stage that consist of remove the noise from given image and segment it to get the object (human body) extracted from background image, then feature extraction to extract the important features needed in the final stage with classification to classify the activities using non-linear SVM classifier.

Fadheela .S

2.1 Noise Removalor De-noising

The presence of noise in image is one of the prime problems in computer vision .Noise appears when the images are taken using fixed camera, two sources imaging device and surveillance camera .Removal of noise from image is prerequisite for processing .Several methods are suggested to handle this difficulty .Among all these methods, the most simplistic and powerful is the standard total variation de-noising introduced by Rudin-Osher Fatemi called ROF model [15].Following the classical form of a regularization approach, we minimize the objective function consisting of data-fidelity combined with regularization term and describe it in a discrete framework. Denoting N as set of nodes of the pixel grid of size h and the image coordinates by x which is a function of (x,y), the following optimization problem need to be solved to obtain smoothed data$\upsilon$from noisy input f,

$$\min_{\upsilon}\left[\left(\sum_{x\in N}\frac{1}{2}\omega(x)(\upsilon(x)-f(x))^2\right)+\lambda R(\upsilon)\right]$$
(1)

Where the $\lambda R(\upsilon)$ is the regularization and $\omega(x)$ is the weight term used to account for the locally varying noise variance. Rescaling of the regularization parameter λ yields $\omega(x)=1$. The regularization parameter $(\lambda > 0)$ controls the amount of smoothing. The regularization term for ROF takes the form,

$$\lambda R(\upsilon) = \sum_{(x,y)}\|\nabla_{\upsilon}(x,y)\| \quad (2)$$
$$= \sum_{(x,y)}\sqrt{([(\upsilon(x+h,y)-\upsilon(x,y))]^2 +}$$
$$(\upsilon(x,y+h)-\upsilon(x,y))^2 \qquad (3)$$

According to this formula, 3 x 3 window is used to cover all given image except corner area that can Stretching and shrinking based on (h) factor to find the averaging values.

2.2 Segmentation

The most significant part in HAR is the segmentation without which features cannot be extracted from an object (part of human body). Conventionally, the segmentation is based on the gradient of the image acting as the stopping term and cause unsatisfactory performance in noisy images[16]. A new technique is proposed that does not consider the edge information but utilizes the difference between the regions inside and outside of the curve .This is one of the most robust and widely used techniques for image segmentation of human body. In the preprocessing stage an object inside the scene is segmented using an energy function defined by,

$$F(c) = \int_{in(c)}|I(x) - C_{in}|^2\,dx + \int_{out(c)}|I(x) - C_{out}|^2\,dx \quad (4)$$

Where$x \in \Omega$ (the image plane)$\subset R^2$, $I:\Omega \rightarrow Z$ is a certain image feature with intensity of color, or texture. Here ,C_{in} and C_{out} are the mean values of image features inside[in(c)] and outside[out(c)] the curve C. By considering image segmentation as a clustering problem the two different segments (clusters) are minimized. This procedure is not efficient for segmented object within inhomogeneous global region but powerful for local minimization such as human body being usually homogenous.

The parameters used inside (local C_{in}) and outside(global C_{out}) of human body and I make this method acceptable in term of still image.

2.3   Chamfer   Matching(complete missing part)

Chamfer matching is a popular technique for finding the set alignment between two edge maps. Let $U=\{u_i\}$ be the best of template and $V=\{v_j\}$ be the set of query image edge maps. Then, the chamfer distance between U and V is defined as the average of distance between each points $u_i \in U$ and its nearest edge in V and is expressed as,

$$d(U,V) = \frac{1}{n}\sum_{u_i\in V}\min_{v_j\in V}|u_i - v_j| \qquad (5)$$

Fadheela .S

Assuming W is a warping function defined on the plain of image and s is a parameter with 2D Euclidean transformation, such that $s \in SE(2), s = (\theta, t_x t_y)$. Here θ is the in-plane rotation angle , t_x and t_y are the translation along x axis y axis, respectively. The action on image points is follows the transformation,

$$W(x, s) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} x + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \qquad (6)$$

This action can check the entire edge of the shape for one template at different angles and for the best alignment parameters $\hat{s} \in SE(2)$ between two edge maps onegets,

$$\hat{s} = \arg \min_{s \in SE(2)} dcm(W(v, s), V)$$
(7)

where $W(U,s)=\{W(u_i,s)\}$.

The matching cost within chamfer matching can be computed efficiently by distance transform image $DT(x) = \min_{v_j} \in V |x - v_j|$ by specifying the distance from each pixel to the nearest edge pixel of V. The evaluation of linear time distance transform of O(n) is obtained via $dcm(U, V) = \frac{1}{n}\sum_{u_{i \in U}} DT(u_i)$.
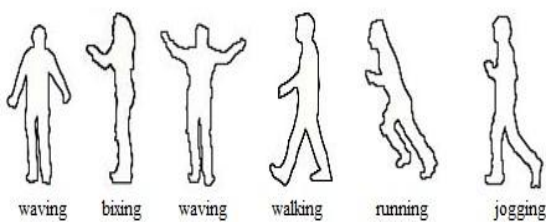

Figure 1 : Full body template for six activities

used by our method.

Let M={$P_1, P_2, \ldots, P_N$} is a template model in which each $P_{i,j} \in \{1, \ldots, N\}$ is a member of the template (or contour) representing the shape of the i-th part of the object. To obtain the template model M, a set of K templates (vector) $\lambda = \{T_1, T_2, \ldots, T_k\}$ considering all possible shapes of the human body are collected as shown in Figure 1. Each template $T_j \in T, j \in \{1, \ldots, k\}$ is centered in a fixed size window and then divided into N parts to create N part templates T_{j1}, $T_{j2}, \ldots, T_{jN}$ as illustrated in Figure 2. We assume that the number of parts (N) and their location $l_i = (x_i, y_i)$, $I \in \{1, \ldots, N\}$ can be determined in advance. For each type of parts (i-th part), the corresponding template set {T_{ji}}, $j \in \{1, \ldots, K\}$ are aligned based on their location li and clustered by using a K-means algorithm. The set pi consists of mean templates signifying the centerscf{T_{ij}}.

Figure 1 depicts the template model M representing shape of the human body. The full human body structure is decomposed into four different parts (N=4) including head-torso or upper, bottom (legs), left and right. The layout of these parts is displayed in Figure 2. The K-means algorithm [20] used 5,8,6 and 6 parts for defining the three main layout left-right, bottom and top which represent the number of clusters.
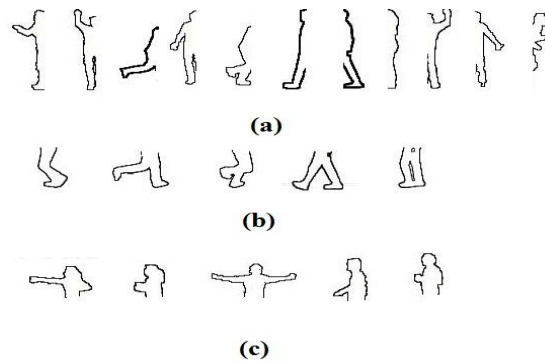

(a)

(b)

(c)

Figure 2 : Parts of template (a) Left-Right part, (b) Bottom part and (c) Upper part.

Fadheela .S

The total number of 5+8+6+6=25 templates (each template is different from another according to its side for example right or left side got six shapes while down side got shapes) is used to cover 5*8*6*6=1440 postures which represent the total number of features in this model. Local appearance features are extracted at the key points (joints). The key points satisfy all the human body and overall activities for the human .Therefore ,the selection of joints as key points makes it reasonable. Each template must have at least four key points to reconstruct the human body.

After defining the template model M, the sets of key points are detected to determine the shapes of the object. A candidate region called detection window at location L=(x,y) is a collection of sets of key points $G(I, M)$ is given by the expression,

$$G(I, M) = \cup_{i=1}^{N} S^{li}(I, P_i) \qquad (8)$$

where ∪ represents the redundant union operator on subset in which the elements of individual sets can be duplicated in the union set,$S^{li}(I, P_i)$. such P is the set of key points representing the shape of the i-th part located at L_i which is obtained from,

$$S^{l}(I, \mathcal{T}) = \{e(t) \in E(I) | t \in T^l\} \qquad (9)$$

where e(t) is the closest point of t in entire point, E(I). Note that, every template T∈Pi is already been sampled at the same number of locations.

This method is better than others in terms of high accuracy when detecting objects included in the still image .However, the calculation is rather time consuming because of the mathematical rigor ,especially if more activities considered online .The approach is considered to be effective for the estimation of the missing parts of the human body widely used in many security and vision applications.

2.4 Features Extraction

In acquiring features we use sampled patch P which is additionally augmented with vector V_p that consisting of associated joints J_K. The functional dependence based on squared distances yields,

$$\mathcal{H}(P) = \frac{1}{|P|}\sum_{p \in P}\|V_{p,k} - M_k\|^2 \qquad (10)$$

where M_k denotes the mean, and V_P consider a vector of joints in one block then can get vector of features $\mathcal{H}(P)$.

The probability of the class, $P(c/L_T)$ and probability of vectors ,$P(V/L_T)$ are stored. The unary potential at location x for the joint K is given by,

$$\Phi_K(J_K(X)) = \sum_{Y \in \Omega}\frac{1}{|T|}\sum_{T \in \mathcal{T}} p(c = k| L_T(p(Y)).p(X - Y|L_T(p(Y))) \qquad (11)$$
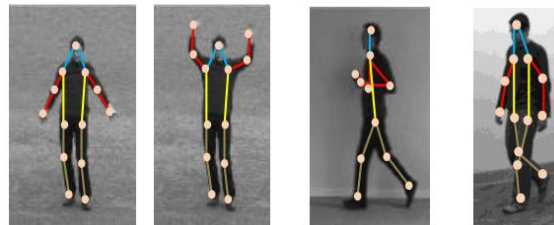
where J∈ {0,1}.

Figure 3 : Simple images from KTM dataset showing stick structure.

Figure 3 depicts the detail mechanism of extracting features from detected object inside the image. Six parts of the body including Torso, Upper leg, Lower leg, Upper arm, Fore-arm and Head make easier to construct the human body for extracting the features according to the structure of pose tree. The tree starts with root in the head joint and continue to the terminal leaf in feet and wrist. The distance among joints in torso and shoulder is calculated to complete entire body features.

Fadheela .S

2.5 classification

In order to identify the activities ,SVM classifier were used to classify each activity according to given vectors of feature extracted from an object, the non-linear classifier clustering the features of such activities in to the groups and these groups tends to be small and far of each other be suitable margin to recognize the activity of specific features.   These features determine the activity belong to which class which are already arrange accordingly as shown in Figure (4)
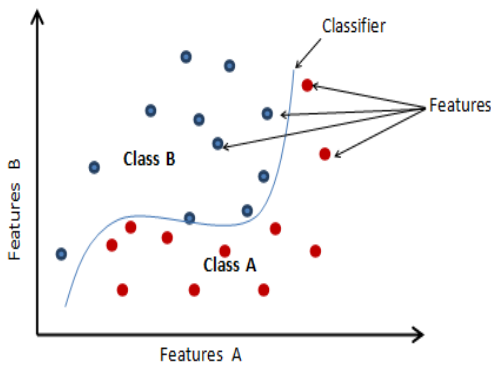

Figure 4: classifier architecture

For more description non-linear formulation of SVM was used to classify w features to put it within class 0 or 1 of n template to all shapes as illustrated bellow:

$$\left[\frac{1}{n}\sum_{i=1}^{n} max\left(0,1 - y_i(w.x_i - b)\right)\right] + \gamma\|w.\|^2$$

Such kernel used with non-linear SVM is y where represented as $y \in \{2^{-2}2^{-1} \dots 2^{1}2^{2}\}$ as class A or B for each pair activitie.

3.   Results and Discussion

Classification of human activities depends on number and types of features extracted from the object .In the proposed method we extract the full features from an image derived from the entire body by detecting the occluded or missing parts. Figure 5 displays the achieved recognition rates of the occluded human body parts.
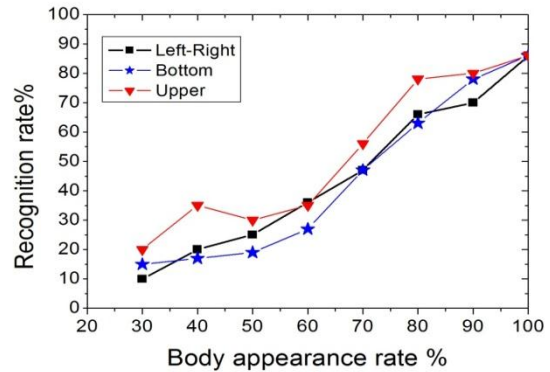

Figure 5: Recognition rates of three body parts with the appearance of each part.

Recognition rate can only be observed beyond 30% because the system cannot recognize less than a certain threshold of extracted features from the part of the body and need to satisfy a minimum number .For instance, head has no features to extract compare to other parts of human body. Meanwhile, the feature types used in this research do not conform to the head of the body which is only 12% of the human body.

Superior recognition rate depends on the positive detection rate and the choice of corresponding features in advanced. In case, the missing part from the body is too much then the positive detection is low and there by harmonizing with correct estimated template become difficult. Figure 5 shows positive recognition with corresponding samples of body part .Here, features are extracted from each part by making a relation between them and the amount of occluded parts .Less recognition rate occurs when the appearance is low enough making it difficult to extract good features needed by the classifier.
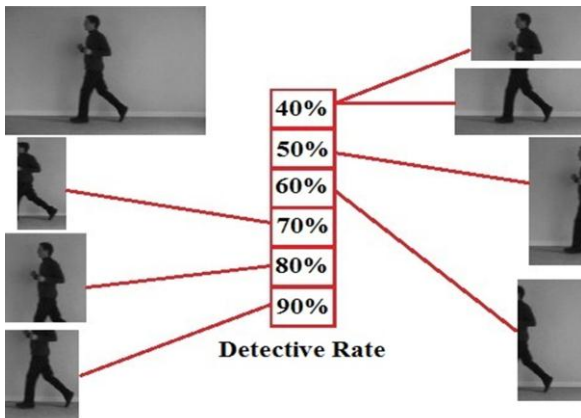
Fadheela .S

Figure 6 : Positive detection rates with corresponding body parts in training system.

Figure 7 : Two different datasets used in our system.

These detective rate are computed experimentally when training the system then become fixed for testing algorithm .There are several images in KTH dataset where the detection of some body parts is very difficult. For example, with arms due to wearing of different clothes hiding the arms and some human position when tilted make them impossible to detect .This imparts additional difficulty in the complete detection of entire body parts using the proposed model. In KTH dataset there are different activities such as walking, jogging, running, waving, boxing and clamping. All these activities are performed via 25 different characters for both sexes with different scenarios including outdoors, indoors, outdoors with scales and outdoors with different clothes. These features are inputted in our proposed non-linear SVM classifier to achieve high-quality and accurate recognition rate.

The recognition rate with proposed method (86 %) is found to be highest with the full extraction of powerful features and the corresponding detection of the human activities .The results obtained by us using the same two datasets as reported earlier display highest recognition rate compare to that of previous researchers. The full color high resolution images from the dataset of INRIA and gray-look low resolution from KTH data set used in this work are shown in Figure 7.

To characterize the activities as much as possible it is important to detect each part of the human body, especially when using some occluded or missing body part we are trying to estimate the whole body according to a given appearance .Detection and estimation of such appearance is significant to built and suggest a suitable template in advance. The features extracted from available incomplete parts of human body renders low recognition rate which need to be enhanced to achieve highest possible recognition rate. The relation between positive detection of objects and features extracted from human body is illustrated in Figure 8. The most important features of human body appearance are observed to be extracted from 55 to 82%.
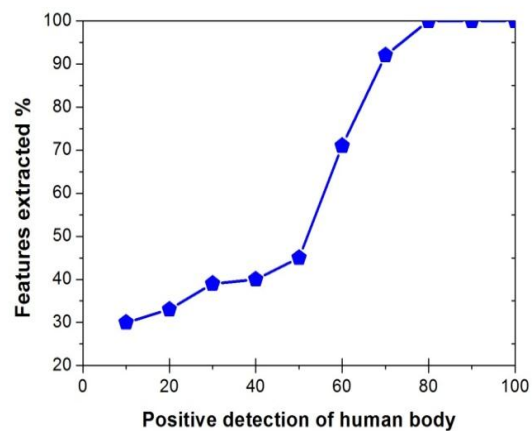

Figure 8 : Relation between feature extraction and human body appearance.

Fadheela .S

The results for the confusion matrix with supervised classification are furnished in Figure 9. Interestingly, for activities such as walking, running and jogging the system is capable of recognizing the features extracted from two parts of the body (arm and leg) and a high recognition rate is accomplished. In contrast, the activities such as boxing and waving achieves relatively lower recognition rate because the features are extracted from one part of the body (arm).

We evaluate the accuracy of our method with seven activities using two different datasets

Figure 9: Confusion matrix showing recognition rate for seven activities.

because some activities are available only in a particular dataset and some belong to another. Training set with INRIA dataset is found to be much more computationally time consuming than with KTH because each image is changed to gray scale followed by noise reduction, segmentation and so on. Meanwhile, training with INRIA dataset consisting of 780 images of (64 × 128) pixels format for each activity takes longer time to learn the machine.

4.   Conclusion

A new method is proposed for recognizing activities in still images for occluded human body or missing part out of the scenes .We design the templates for some activities with different sides to complete and estimate occluded body parts for extracting powerful features from stick structure which are divided into six parts

according to the pose tree of joints of human body. The classifier is used in three main stages including preprocessing (foreground extraction), segmentation (background subtraction) to extract useful features from object and sort out these features by the classifier (classification). The process of de-noising and chamfer matching allows us to obtain high-quality and accurate features. Our method achieves the highest recognition rate as much as 86% for seven activities including running, walking, boxing, jogging, waving, capping and jumping. This method despite time expensive is capable of extracting features from partially appearance human body parts and fully estimated body to classify the activities using non-linear SVM classifier using two public datasets KTH and INRIA .The model is simulated using MATLAB programming for several images in high pixels format. We assert that the appropriate implementation using suitable human computer interfaces critically depend on such image analyses useful for widespread applications from surveillance to security systems to automation.

References

[1] Matthias D, Juergen G, Christian L, Luc V. 2013 .Human Pose Estimation using Body Parts Dependent Joint Regressors. In: IEEE 2013 Computer Vision and Pattern Recognition; 23-28; Portland, :IEEE.pp.2041 – 2048.

[2] FischlerM ,Elschlager R.The 1973. representation and matching of pictorial structures. IEEE ;22:67-92.

[3] Pedro F, Daniel P. 2005 .Pictorial Structures for Object Recognition.ACM DL; 61:55-79.

[4] Aggarwal J, Ryoo M. 2011 .Human activity analysis: A review. ACM DL; 43:157-202.

[5] Vincent D , Ivan L , Josef S. 2010. Recognizing human actions in still images: a study of bag-of-features and part-based representations. CiteSeer;22:50-61.

Fadheela .S

[6] Hueihan J, Juergen G, Silvia Z, Cordelia S, Michael J. 2013 .Towards understanding action recognition. In : IEEE 2013 Computer Society ; 15-20;USA,:IEEE.pp.3192-3199.

[7] Prest A1, Ferrari V, Schmid C. 2013 .Explicit modeling of human-object interactions in realistic videos.IEEE Trans Pattern Anal Mach Intell;35:835-848.

[8] BO W, RAM N. 2007. Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors. Springer International Journal of Computer Vision;75: 247-266.

[9] Sung J, Mark S.Heel 2012. strike detection based on human walking movement for surveillance analysis. ELSEVIAR Pattern Recognition Letters; 34: 895–902.

[10] Trung T, Phong V, Bac L.2009. Detecting probable regions of humans in still images using raw edges.In :IEEE 2009 Knowledge and Systems Engineering; 13-17 October; Hanoi,:IEEE.pp. 73 – 80.

[11] Mehmet A, Alper B, Mehmet E.2013. A novel fuzzy filter for speckle noise removal. Turk J ElecEng& Comp Sci;10.3906:1210-24.

[12] Habil K. 2013.Online feature selection and classification with incomplete data. Turk J ElecEng& Comp Sci; 10.3906:1301-181.

[13] Halime B. and Yüksel Ö. 2013 .A new method for segmentation of microscopic images on activated sludge. Turk J ElecEng& Comp Sci; 10.3906:1307-9.

[14] Michalis V, Vasileios K, Christophoros N, Ioannis A. 2014.Matching Mixtures of Trajectories for Human Action Recognition. Computer Vision and Image Understanding; 19: 27-40.

[15] Marcin G, Christian T, Reinhard K,2012. Andreas K.Time-of-Flight and Depth Imaging. Springer; 5: 319-324.

[16] Muhammad H, Adil M, Seok-Won L.2013. Active Contours Level Set Based Still Human Body Segmentation from Depth Images For Video-based Activity Recognition.KSII Transactions On Internet And Information Systems; 7: 2839-2852.

[17] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. 2002 . An efficient k-means clustering algorithm: Analysis and implementation. IEEE transactions on pattern analysis and machine intelligence, 24(7), 881-892.

Fadheela .S

نموذج للتعرف على الانشطة البشرية باستخدام صور مغلقة جزئيا

فضيلة صبري ابو الماش
دائرة البعثات والعلاقات الثقافية

المستخلص :

ان الحاجة الى انظمة التعرف على الانشطة البشرية (HAR) يتم من خلال الصور الثابتة التي ازدادت في الآونة الاخيرة وما يقابلها من تصنيفات للمواصفات واستخلاصها وتصنيفها بالمصنفات القياسية المتوفرة. انظمة التعرف على الانشطة البشرية التي تهتم بالجزء المفقود من الصورة يحوي على تحدي كبير وذلك لقلة التعامل مع الانظمة التي تعتمد على الصور الثابتة اما الصور المتحركة والفيديو فهي شائعة الاستعمال. لذلك التحدي في الصور الثابتة كون المعلومات المسبقة للصورة لا تتوفر ويلزم التنبؤ بها.

اهمية البحث تأتي من المصالح الرئيسية التي تعتمد على كامرات المراقبة في الشركات الامنية والمصانع. ففي هذا البحث نقترح نموذج جديدا لتمييز الانشطة البشرية للجزء المفقود من الجسم في صورة ثابتة بعد التنبؤ بالجزء المفقود ثم استخراج المواصفات للشكل وبعد تصنيفه نحصل على دقة مناسبة. اشكال معينة بنيت لإتمام الاجسام في الصورة الذي يلزم لاستخراج المواصفات الكلية للجسم بسبب انه لايمكن للنظام التنبؤ الا في حالة اكمال الجسم. مشكلة التصنيف تم معالجتها عن طريق مصنف قياسي معروف (SVM). بعض الاجراءات اتبعت للصور قبل المعالجة وذلك عن طريق ازالة الضوضاء من الصورة وكذلك في نظامنا تم استعمال ١٢٠٠ صورة ثابتة بدقة (١٢٨*٦٤) بكسل من قاعدة بيانات قياسية (INRIA, KTH) . معدل الدقة المنجز هنا تقريبا ٨٦% مع سبعة فعاليات بشرية هي: الركض، الجري، المشي، القفز، النشر، الملاكمة، حني اليد. المواصفات التي تم الحصول عليها تعتبر واعدة بالنسبة الى الانظمة ذات الصور الثابتة.